Applicant : Short et al. (as amended) Attorney's Docket No.: 56446-20001.20/-
Serial No. : 09/884,889 002006/D1100-4US
Filed : June 19, 2001
Page : 2 of 29

Amendment to the Specification:

Please amend the specification as follows.

Please replace the paragraph lines 16 to 19, on page 2, with the following amended paragraph:

Another aspect of the invention is an isolated nucleic acid encoding a polypeptide or a [[funcional]] functional fragment thereof, having a sequence as set forth in Group B amino acid sequences, and sequences substantially identical thereto.

Please replace the paragraph of line 14, page 6, to line 2 or page 7, with the following amended paragraph:

The term "polypeptide" as used herein, refers to amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain modified amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also a given polypeptide may have many types of modifications. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a [[phosphytidylinositol]] phosphatidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristolyation, oxidation, pegylation [[pergylation]], proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to protein such as arginylation. (See Creighton, T.E., Proteins – Structure and Molecular Properties 2nd Ed., W.H. Freeman and Company, New York (1993); Posttranslational Covalent Modification of Proteins, B.C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)).

Please replace the paragraph of line page 10, lines 1 to 17, with the following amended paragraph:

Additionally a "substantially identical" amino acid sequence is a sequence that differs from a reference sequence by one or more conservative or non-conservative amino acid substitutions, deletions, or insertions, particularly when such a substitution occurs at a site that is not the active site of the molecule, and provided that the polypeptide essentially retains its functional properties. A conservative amino acid substitution, for example, substitutes one amino acid for another of the same class (e.g., substitution of one hydrophobic amino acid, such as [[isoleucine]] isoleucine, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid or glutamine for asparagine). One or more amino acids can be deleted, for example, from an catalase polypeptide, resulting in modification of the structure of the polypeptide, without significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for catalase biological activity can be removed. Modified polypeptide sequences of the invention can be assayed for catalase biological activity by any number of methods, including contacting the modified polypeptide sequence with an catalase substrate and determining whether the modified polypeptide decreases the amount of specific substrate in the assay or increases the bioproducts of the enzymatic reaction of a functional catalase polypeptide with the substrate.

Please replace the paragraph lines 19 to 25, on page 11, with the following amended paragraph:

The term "variant" refers to polynucleotides or polypeptides of the invention modified at one or more base pairs, codons, introns, exons, or amino acid residues (respectively) yet still retain the biological activity of a catalase of the invention. Variants can be produced by any number of means included methods such as, for example, error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, in vivo mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble

mutagenesis, site-specific mutagenesis, gene reassembly, [[(GSSM)]] GSSM™ and any combination thereof.

Please replace the paragraph lines 6 to 10, on page 29, with the following amended paragraph:

Optionally, the method comprises the additional step of screening the library members of the shuffled pool to identify individual shuffled library members having the ability to bind or otherwise interact, or catalyze a particular reaction (e.g., such as catalytic domain of an enzyme) with a predetermined macromolecule, such as for example a proteinaceous receptor, an oligosaccharide, [[viron]] virion, or other predetermined compound or structure.

Please replace the paragraph of line 17, on page 29, to line 6 of page 30, with the following amended paragraph:

In another aspect, it is envisioned that prior to or during recombination or reassortment, polynucleotides generated by the method of the invention can be subjected to agents or processes which promote the introduction of mutations into the original polynucleotides. The introduction of such mutations would increase the diversity of resulting hybrid polynucleotides and polypeptides encoded therefrom. The agents or processes which promote mutagenesis can include, but are not limited to: (+)-CC-1065, or a synthetic analog such as (+)-CC-1065-(N3-Adenine, see Sun and Hurley, 1992); an [[N-acelylated]] N-acetylated or deacetylated 4'-fluro-4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see, for example, van de Poll et al., 1992); or a N-acetylated or deacetylated 4-aminobiphenyl adduct capable of inhibiting DNA synthesis (see also, van de Poll et al., 1992, pp. 751-758); trivalent chromium, a trivalent chromium salt, a polycyclic aromatic hydrocarbon ("PAH") DNA adduct capable of inhibiting DNA replication, such as 7-bromomethyl-benz[a]anthracene ("BMA"), tris(2,3-dibromopropyl)phosphate ("Tris-BP"), 1,2-dibromo-3-chloropropane ("DBCP"), 2-bromoacrolein (2BA), benzo[a]pyrene-7,8-dihydrodiol-9-10-epoxide ("BPDE"), a platinum(II) halogen salt, N-hydroxy-2-amino-3-methylimidazo[4,5-f]-quinoline ("N-hydroxy-IQ"), and N-hydroxy-2-amino-1-methyl-6-phenylimidazo[4,5-f]-pyridine ("N-hydroxy-PhIP"). Especially preferred means for slowing or halting PCR amplification consist of UV light (+)-CC-1065 and

Applicant : Short et al. (as amended)   Attorney's Docket No.: 56446-20001.20/-
Serial No. : 09/884,889   002006/D1100-4US
Filed   : June 19, 2001
Page   : 5 of 29

(+)-CC-1065-(N3-Adenine). Particularly encompassed means are DNA adducts or polynucleotides comprising the DNA adducts from the polynucleotides or polynucleotides pool, which can be released or removed by a process including heating the solution comprising the polynucleotides prior to further processing.

Please replace the paragraph lines 12 to 21, on page 30, with the following amended paragraph:

The invention also provides for the use of proprietary codon primers (containing a degenerate N,N,N sequence) to introduce point mutations into a polynucleotide, so as to generate a set of progeny polypeptides in which a full range of single amino acid substitutions is represented at each amino acid position ([[gene site saturated mutagenesis (GSSM)]] Gene Site Saturation Mutagenesis™ (GSSM™)). The oligos used are comprised contiguously of a first homologous sequence, a degenerate N,N,N sequence, and preferably but not necessarily a second homologous sequence. The downstream progeny translational products from the use of such oligos include all possible amino acid changes at each amino acid site along the polypeptide, because the degeneracy of the N,N,N sequence includes codons for all 20 amino acids.

Please replace the paragraph line 14, on page 37, to line 2, page 38, with the following amended paragraph:

Alternatively, more than one probe (at least one of which is capable of specifically hybridizing to any complementary sequences which are present in the nucleic acid sample), may be used in an amplification reaction to determine whether the sample contains an organism containing a nucleic acid sequence of the invention (e.g., an organism from which the nucleic acid was isolated). Typically, the probes comprise oligonucleotides. In one embodiment, the amplification reaction may comprise a PCR reaction. PCR protocols are described in Ausubel and Sambrook, supra. Alternatively, the amplification may comprise a ligase chain reaction, 3SR, or strand displacement reaction. (See Barany, F., "The Ligase Chain Reaction in a PCR World", *PCR Methods and Applications* 1:5-16, 1991; E. Fahy *et al.*, "Self-sustained Sequence Replication (3SR): An Isothermal Transcription-based Amplification System Alternative to PCR", *PCR Methods and Applications* 1:25-33, 1991; and Walker G.T. *et al.*, "Strand Displacement

Amplification-an Isothermal *in vitro* DNA Amplification Technique", *Nucleic Acid Research*
20:1691-1696, 1992, the disclosures of which are incorporated herein by reference in their
entireties). In such procedures, the nucleic acids in the sample are contacted with the probes, the
amplification reaction is performed, and any resulting amplification product is detected. The
amplification product may be detected by performing gel electrophoresis on the reaction products
and staining the gel with an [[interculator]] intercalator such as ethidium bromide. Alternatively,
one or more of the probes may be labeled with a radioactive isotope and the presence of a
radioactive amplification product may be detected by autoradiography after gel electrophoresis.

Please replace the paragraph of line 3, page 60, to line 10, page 61 with the
following amended paragraph:

A "comparison window", as used herein, includes reference to a segment of any one
of the number of contiguous positions selected from the group consisting of from 20 to 600, usually
about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared
to a reference sequence of the same number of contiguous positions after the two sequences are
optimally aligned. Methods of alignment of sequence for comparison are well-known in the art.
Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology
algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482, 1981, by the homology alignment
algorithm of Needleman & Wunsch, *J. Mol. Biol* 48:443, 1970, by the search for similarity method
of person & Lipman, *Proc. Nat'l. Acad. Sci.* USA 85:2444, 1988, by computerized implementations
of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software
Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by manual alignment and
visual inspection. Other algorithms for determining homology or identity include, for example, in
addition to a BLAST program (Basic Local Alignment Search Tool at the National Center for
Biological Information), ALIGN, AMAS (Analysis of Multiply Aligned Sequences), AMPS
(Protein Multiple Sequence Alignment), ASSET (Aligned Segment Statistical Evaluation Tool),
BANDS, BESTSCOR, BIOSCAN (Biological Sequence Comparative Analysis Node), BLIMPS
(BLocks IMProved Searcher), FASTA, Intervals & Points, BMB, CLUSTAL V, CLUSTAL W,
CONSENSUS, LCONSENSUS, WCONSENSUS, Smith-Waterman algorithm, DARWIN, Las
Vegas algorithm, FNAT (Forced Nucleotide Alignment Tool), Framealign, Framesearch,

DYNAMIC, FILTER, FSAP (Fristensky Sequence Analysis Package), GAP (Global Alignment

Program), GENAL, GIBBS, GenQuest, ISSC (Sensitive Sequence Comparison), LALIGN

(Local Sequence Alignment), LCP (Local Content Program), MACAW (Multiple Alignment

Construction & Analysis Workbench), MAP (Multiple Alignment Program), MBLKP, MBLKN,

PIMA (Pattern-Induced Multi-sequence Alignment), SAGA (Sequence Alignment by Genetic

Algorithm) and WHAT-IF. Such alignment programs can also be used to screen genome

databases to identify polynucleotide sequences having substantially identical sequences. A

number of genome databases are available, for example, a substantial portion of the human

genome is available as part of the Human Genome Sequencing Project (J. Roach,

~~http://weber.u.Washington.edu/~roach/human_genome_progress 2.html) (Gibbs, 1995)~~. At least

twenty-one other genomes have already been sequenced, including, for example, *M. genitalium*

(Fraser et al., 1995), *M. jannaschii* (Bult et al., 1996), *H. influenzae* (Fleischmann et al., 1995), *E.*

coli (Blattner et al., 1997), and yeast (*S. cerevisiae*) (Mewes et al., 1997), and *D. melanogaster*

(Adams et al., 2000). Significant progress has also been made in sequencing the genomes of model

organism, such as mouse, *C. elegans*, and *Arabadopsis sp.* Several databases containing genomic

information annotated with some functional information are maintained by different organizations,

and are accessible via the internet, ~~for example, http://www.tigr.org/tdb;~~

~~http://www.genetics.wisc.edu; http://genome-www.stanford.edu/~ball; http://hiv-web.lanl.gov;~~

~~http://www.ncbi.nlm.nih.gov; http://www.ebi.ac.uk; http://Pasteur.fr/other/biology; and http://~~

~~www.genome.wi.mit.edu~~.

Please replace the paragraph of line 11, page 61, to line 2, page 62 with the

following amended paragraph:

One example of a useful algorithm is BLAST and BLAST 2.0 algorithms, which are

described in Altschul et al., *Nuc. Acids Res.* 25:3389-3402, 1977, and Altschul et al., *J. Mol. Biol.*

215:403-410, 1990, respectively. Software for performing BLAST analyses is publicly available

through the National Center for Biotechnology Information [[(http://www.ncbi.nlm.nih.gov/)]].

This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short

words of length W in the query sequence, which either match or satisfy some positive-valued

threshold score T when aligned with a word of the same length in a database sequence. T is referred

to as the neighborhood word score threshold (Altschul et al., supra). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4 and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength of 3, and expectations (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff, *Proc. Natl. Acad. Sci.* USA 89:10915, 1989) alignments (B) of 50, expectation (E) of 10, M=5, N= -4, and a comparison of both strands.

Please replace the paragraph of line 25, page 62, to line 5, page 63 with the following amended paragraph:

The BLAST programs identify homologous sequences by identifying similar segments, which are referred to herein as "high-scoring segment pairs," between a query amino or nucleic acid sequence and a test sequence which is preferably obtained from a protein or nucleic acid sequence database. High-scoring segment pairs are preferably identified (i.e., aligned) by means of a scoring matrix, many of which are known in the art. Preferably, the scoring matrix used is the BLOSUM62 matrix (Gonnet et al., *Science* 256:1443-1445, 1992; Henikoff and Henikoff, *Proteins* 17:49-61, 1993). Less preferably, the PAM or PAM250 matrices may also be used (see, e.g., Schwartz and Dayhoff, eds., 1978, *Matrices for Detecting Distance Relationships: Atlas of Protein Sequence and Structure*, Washington: National Biomedical Research Foundation). BLAST programs are accessible through the U.S. National Library of Medicine, [[e.g., at www.ncbi.nlm.nih.gov]].

Please replace the paragraph of line 27, page 67, to line 9, page 68 with the following amended paragraph:

Figure 5 is a flow diagram illustrating one embodiment of an identifier process 300 for detecting the presence of a feature in a sequence. The process 300 begins at a start state 302 and then moves to a state 304 wherein a first sequence that is to be checked for features is stored to a memory 115 in the computer system 100. The process 300 then moves to a state 306 wherein a database of sequence features is opened. Such a database would include a list of each feature's attributes along with the name of the feature. For example, a feature name could be "Initiation Codon" and the attribute would be "ATG". Another example would be the feature name "TAATAA Box" and the feature attribute would be "TAATAA". An example of such a database is produced by the University of Wisconsin Genetics Computer Group [[(www.gcg.com)]]. Alternatively, the features may be structural polypeptide motifs such as alpha helices, beta sheets, or functional polypeptide motifs such as enzymatic active sites, helix-turn-helix motifs or other motifs known to those skilled in the art.